

항만 안전사고 데이터를 활용한 사전학습 언어모델 기반 안전사고 예측 모델 개발

이유진, 문기영, 박재현, 김도현, 이상덕*

성균관대학교, 한국철도기술연구원

leeyoujin225@g.skku.edu, mkyoung21@krri.re.kr, zephyr@krri.re.kr, dohyunkim@krri.re.kr,
*sdlee@krri.re.kr

Development of Safety Accident Prediction Model Based on Pre-Learning Language Model Using Port Safety Accident Data

Yujin Lee¹, Ki-Yeong Moon², Dohyunkim², Jaehyun Park², Sangduck Lee^{2*}

¹Sungkyunkwan Univ., ²Korea Railroad Research Institute.

요약

최근 항만 하역 산업재해는 감소하고 있지만, 여전히 타 산업에 비하여 사망자 등 중대재해의 발생 비율이 높다. 항만 안전사고 예방을 위해 방대한 양의 비정형 데이터가 존재하지만, 해당 자료는 사례집, 교육자료, 뉴스 등 다양한 형태로 되어 있어 정작 안전관리자나 작업자들이 이를 모두 확인하기 어렵다. 따라서 본 논문에서는 비정형으로 이루어진 항만 안전사고 데이터를 자동으로 분석하기 위해 사전학습 언어모델을 활용하여 안전사고별 사고 유형을 예측하였다. 이를 위해 사고 당시 상황을 기록한 보고서를 수집해 사고 유형별 분류 데이터 셋을 구축하고 SKT-Brain에서 공개한 KoBERT의 사전학습 모델을 파인튜닝하여 사고 유형을 예측하였다. 본 연구에서 제안한 안전사고 예측 모델은 평가 결과, Accuracy 0.91의 안전사고 예측 정확도를 가진다.

1. 서론

산업규모가 점차 확대됨에 따라 작업이 혼잡해지고 대형장비가 증가하여 작업자의 안전사고위험이 증가하고 있다. 항만물류협회에 따르면 항만하역 산업재해는 과거 10년전에 비해서 재해자수는 44.1% 감소하였지만, 여전히 타 산업에 비하여 사망자 등 중대재해의 발생비율은 높은 상태이다[1]. 기존의 항만 산업에서는 안전관리자의 지식과 경험에 기반한 안전사고 관리 및 대응을 하기 때문에 다양한 안전사고 발생 원인을 사람이 모두 분석하고 파악하기 어려웠다. 특히, 현장 경험이 부족한 신입 안전관리자는 더욱 어려움을 겪는다. 또한, 항만 안전사고 예방을 위해 항만물류협회는 '항만하역재해통계 및 사례'와 같이 매년 발생한 중경상 및 사망 사고에 대한 현황과 원인 등을 통계적으로 분석하여 제공하고 있다. 이처럼 안전사고 예방에 관한 방대한 양의 비정형 데이터가 존재하지만, 해당 자료는 사례집, 교육자료, 뉴스 등 다양한 형태로 되어 있어 안전관리자나 작업자들이 이를 모두 확인하기 어렵다.

최근 인공지능, 빅데이터, IoT 등 4차 산업혁명 기술이 발전하면서 안전관리 분야에서도 관련 데이터를 수집 및 분석하여 유의미한 정보만을 사용자에게 제공해주는 연구가 증가하고 있다. 특히, 안전사고가 발생할 때마다 사고의 상황을 상세히 기록하는 안전사고 보고서를 활용하여 사고를 예측하거나, 사고 위험 요인을 도출하는 연구가 존재한다[2]. 이처럼 자연어를 기계로 처리하기 위해 Word2Vec, Glove, Fasttext 등과 같이 단어 임베딩 모델이 존재한다. 하지만, 해당 모델들의 경우 단어마다 고유한 벡

터값을 가지는 static embedding 방식이라는 한계가 있다. 따라서 최근에는 이러한 단점을 극복하고 문맥을 이해하여 동적으로(dynamic) 단어를 벡터로 변환하는 딥러닝 기반 자연어 처리 모델이 주로 사용되고 있다.

따라서 본 논문에서는 항만 안전사고와 관련한 방대한 양의 비정형 데이터를 분석하여 안전관리자나 작업자에게 유의미한 정보를 제공하기 위해 딥러닝 기반 자연어 처리 모델을 활용하였다. 항만 안전사고 관련 비정형 데이터를 수집하여 데이터 셋을 구축하고 이를 딥러닝 기반 자연어 처리 모델인 KoBERT[3]로 파인튜닝하여 사고 상황별 사고 예측에 활용하였다. 2장에서는 관련 연구를 소개하고, 3장에서는 항만 안전사고 예측을 위한 사전학습 언어모델을 이용한 실험을 보여준다. 4장에서는 결론 및 향후 연구 방향에 대해 논의하였다.

2. 항만 안전사고 예측 모델

2.1. 데이터 셋

본 연구에서는 항만 안전사고 예측을 위해 한국항만물류협회에서 발간한 재해통계 및 사례집[1]과 산업 현장 내 사고 사례 공유 사이트인 UlsanSafety에서 총 1796개의 사고 사례를 수집하였다. 사례집에서 제시하는 사고 유형 종류를 참고하여 9종의 사고 유형을 선정했고 각 사고 유형에 대한 정의는 한국산업안전공단에서 2006년에 발간한 산업재해 기록·분류에 관한 지침을 참고했다. 그 결과, 사고 유형별 정의는 표 1과 같으며, 1796개의 사고 사례에 대해 정의된 사고 유형으로 라벨링하여 데이터 셋을 구축

하였다.

표 1 항만 안전사고 유형 9종 정의

	사고 유형	정의
1	추락	떨어짐 (높이가 있는 곳에서 사람이 떨어짐)
2	전도	넘어짐(사람이 미끄러지거나 넘어짐), 깔림(물체의 쓰러짐이나 뒤집힘)
3	충돌	부딪힘(물체에 부딪힘)
4	낙하	맞음(떨어진 물체에 맞음)
5	비래	맞음(날아온 물체에 맞음)
6	붕괴	무너짐(건축물이나 쌓인 물체가 무너짐)
7	협착	끼임(기계설비에 끼이거나 감김)
8	유해물	유해·위험물질에 노출·접촉 또는 흡입하였거나 독성동물에 쏘이거나 물린 경우
9	폭발	건축물, 용기내 또는 대기중에서 물질의 화학적, 물리적 변화가 급격히 진행되어 열, 폭발음, 폭발압이 동반하여 발생하는 경우

2.2. 제안 모델 학습 및 평가 결과

본 연구에서 SKT-Brain에서 공개한 KoBERT[3]의 사전학습 모델에 구축한 항만 안전사고 유형 분류 데이터 셋을 학습시켜 안전사고 예측을 위한 파인튜닝을 수행하였다. 안전사고 사례의 사고 유형을 분류하기 위해 KoBERT의 문장 분류 기능을 활용하여 사고 사례별 사고 유형을 분류하였다. 모델 학습을 위해 2.1절에서 구축한 항만 안전사고 유형 분류 데이터 셋을 사용하였다. 총 데이터의 수는 1,756개이며, 이를 8대 2로 나누어 훈련과 검증을 진행하였다. 학습에 사용된 하이퍼파라미터는 표2와 같다.

표 2 모델 학습을 위한 하이퍼파라미터

하이퍼파라미터	내용
배치 사이즈 (batch size)	16
에폭 수 (epochs)	5
학습률 (learning rate)	5e-5
옵티마이저 (optimizer)	AdamW
손실 함수(loss function)	Cross Entropy
최대 토큰 길이 (max token length)	256

KoBERT를 구축한 항만 안전사고 유형 분류 데이터셋으로 학습된 모델의 성능을 분석하기 위해 평가지표로 정확도(Accuracy)를 사용하였다. 정확도(Accuracy)는 수식(1)과 같으며, 분류 성능을 측정하는 평가지표로 가장 많이 활용된다. 모델 평가 결과, 항만 안전사고 예측 정확도(Accuracy)는 0.91이다.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{(\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative})} \quad (1)$$

또한, 학습에 사용되지 않은 안전사고 사례를 입력으로 넣은 후 안전사고 예측 결과를 테스트해본 결과, 표 3과 같이 안전사고를 예측하는 것을 볼 수 있다. 작성한 충돌, 붕괴, 추락, 협착, 낙하, 유해물 관련 예시 이외에도 항만 안전사고 유형 9종에 대한 사고 예측을 원활히 수행함을 확인하였다.

표 3 항만 안전사고 유형 예측 결과

Input Sentence	Predicted Result
2012년 12월 10일 인천항 60선석에서 H빔 이동작업 중 지게차 조작미숙으로 화물이 한 쪽으로 쏠려 가슴 부위가 핸들에 부딪혀 부상	충돌
2018년 1월 24일 20시경 제주항 ○○부두 야적장에서 5톤차량에 적재된 화물을 지게차로 하차작업 중 차량에 적재된 폴컨테이너 화물이 갑자기 붕괴되어 재해자가 이를 막으려고 하다 화물에 깔려 사망한 사고	붕괴
2012년 11월 28일 인천항 1부두에서 화물상·하차 작업 중 백에 연결된 로프를 잡고 하차 중 로프가 끊어져 바닥으로 떨어짐. 피재해자가 어깨에 경상을 입음	추락
2021.2.8.(월) 00제철소 내 원료하역 부두에서 컨베이어 롤러 교체작업 중 원료 하역기가 이동하면서 컨베이어와의 사이에 끼임.	협착
2013년 2월 7일 인천항 1부두 14야적장에서 야적 되어있는 소금 백을 상차하던 중 떨어져 오른팔 부상	낙하
전남 여수시의 화학공장에서 화학물질 배관 변경과정에서 밸브 잠금 불량으로 사이클로렉산 화학물질이 누출되는 사고가 발생하였다. 이 사고로 12명이 흡입하여 경상을 입었다	유해물

4. 결론 및 향후 연구

본 연구에서는 KoBERT 모델을 활용하여 항만 안전사고 예측 모델을 개발하였다. 항만 안전사고 유형 분류 데이터 셋을 구축하였고 이를 통해 KoBERT 모델을 파인튜닝하여 안전사고 사례가 주어지면 사고 유형을 예측하도록 하였다. 모델 평가 결과, Accuracy가 0.91인 항만 안전사고 예측 정확도를 얻었다. 향후 연구로는 사고 예측뿐만 아니라 문장으로 이루어진 안전사고 사례 내 중요 정보를 추출할 수 있는 사전학습 언어모델을 개발하고자 한다.

ACKNOWLEDGMENT

이 연구는 2023년도 해양수산부 및 해양수산과학기술진흥원(KIMST) 연구비 지원에 의한 연구임(1525013138, 빅데이터·인공지능 기반 물류연계 최적화 기술 개발 사업)

참 고 문 헌

[1] 한국항만물류협회, “항만하역재해통계및사례”, 2013년 - 2020년 각호

[2] Cho, et al. “Development of Machine Learning-based Construction Accident Prediction Model Using Structured and Unstructured Data of Construction Sites” Korean Society of Civil Engineers, vol.42, no.1, pp.127-134,

[3] SKT. (2021). KoBERT (Version 0.2.3) [Source code]. <https://sktelecom.github.io/project/kobert>